

**Statistica Sinica Preprint No: SS-2016-0163R2**

<b>Title</b>	Sequential Pareto Minimization of Physical Systems Using Calibrated Computer Simulators
<b>Manuscript ID</b>	SS-2016.0163
<b>URL</b>	<a href="http://www.stat.sinica.edu.tw/statistica/">http://www.stat.sinica.edu.tw/statistica/</a>
<b>DOI</b>	10.5705/ss.202016.0163
<b>Complete List of Authors</b>	Po-Hsu Chen Thomas Santner and Angela Dean
<b>Corresponding Author</b>	Po-Hsu Chen
<b>E-mail</b>	chen.2719@osu.edu, allen19850308@gmail.com
Notice: Accepted version subject to English editing.	

# SEQUENTIAL PARETO MINIMIZATION OF PHYSICAL SYSTEMS USING CALIBRATED COMPUTER SIMULATORS

Po-Hsu Allen Chen, Thomas J. Santner and Angela M. Dean

*The Ohio State University*

*Abstract:* This paper proposes a sequential design methodology for a combined physical system and computer simulator experiment having multiple outputs, when the goal is to find the Pareto Front and Set of the *means of the physical system outputs*; the methodology is based on a statistically-calibrated simulator. In this paper, the simulator is a computer implementation of a deterministic mathematical model of the physical system; it contains the same set of control(able) inputs as those used to represent the physical system, plus additional calibration inputs for adjusting the simulator output to better mimic the mean of the physical system. A minimax fitness function is proposed for guiding the sequential search for new vectors of control input settings when additional observations on the *physical system* are to be taken. Based on a Bayesian calibrated model, the update step maximizes the posterior expected minimax fitness function over untried control inputs. When additional runs of the *simulator* are to be taken, the control input settings are chosen as above; then calibration input settings are selected to minimize the sum, over the set of predicted output means, of the posterior mean squared prediction errors. Using the Hypervolume Indicator function to assess Pareto Front accuracy, the performance of the sequential procedure is evaluated using analytic test functions from the multiple-objective optimization literature.

*Key words and phrases:* Combined physical and simulator experiment; Computer experiment; Multiobjective optimization.

## 1 Introduction

Many applications require the simultaneous optimization of multiple objective functions. Examples include those of Forrester, Sobester and Keane (2008) who considered the engineering design of a cantilever beam subject to fabrication constraints; Leatherman, Guo, Gilbert, Hutchinson, Maher and Santner (2014) who studied tissue engineering of the material properties of a meniscal substitute

to provide both small mean and low variability in the peak contact stresses of the underlying cartilage; and Picheny (2015) who considered the optimization of a set of measures of the light intercepted by a tree as a function of breedable tree traits. The references in all three of these articles provide additional subject matter details.

Experiments using computer simulators have been employed to study physical systems when the relationship between system inputs and outputs can be modeled mathematically but experiments using the physical system are infeasible or costly to run. When only a limited number of runs of the simulator is feasible, Sacks, Welch, Mitchell and Wynn (1989) proposed a methodology for predicting simulator output by modeling it as a realization of a Gaussian process (GP). However, computer simulators often have the defect that they provide biased output for the physical system due to the simplified biology or physics used in the mathematical model that they implement. If both physical system observations and simulator output are available, Bayesian calibration can be used to align simulator output to that of the target physical system by constructing a bias-corrected simulator.

In this paper, we are concerned with the simultaneous optimization (minimization, hereafter) of multiple objective functions. Because conflicting objective functions need not have a common optimizer, the minimization is in the sense of *Pareto minimization*, a method which identifies a set of compromise solutions. The inputs of these compromise solutions are termed the *Pareto Set* for the problem and the corresponding outputs constitute the *Pareto Front*. Pareto minimization of simulators based on surrogate predictors has become an active research area (e.g., Keane, 2006; Picheny, 2015; Svenson and Santner, 2016).

This paper introduces a sequential design and analysis method based on a Bayesian calibrated simulator to estimate the Pareto Front of the mean responses of a multiple output physical system. The inputs to the simulator include the control inputs for the physical system plus additional inputs, called *calibration parameters*, that are unknown or difficult to measure in the physical system but which are needed in the mathematical description of the input/output relationship. In the sequential procedure, when an additional observation is to be added from the physical system, a minimax fitness function is proposed to guide the

update step for the selection of the next control input setting. Alternatively, if an additional run is to be made using the computer simulator, the update step selects the control input settings for the next simulator run as above and selects the corresponding calibration input settings to minimize the sum of the posterior mean square prediction errors.

Section 2 describes the calibration model used in this paper, while Section 3 defines the Pareto Front and Pareto Set. The proposed sequential methodology is presented in detail in Section 4. The methodology is illustrated using analytic functions with different output dimensions in Section 5. Section 6 provides a summary and discussion.

## 2 Bayesian Calibration

This paper uses a Bayesian analysis of a statistical calibration model to predict the mean responses used in the proposed design and analysis methodology. The calibration model is due to Kennedy and O’Hagan (2001) (the “KO model”). The Bayesian analysis of this model is provided by Higdon et al. (2004) and Higdon et al. (2008). The KO model assumes that data from both a physical system and a computer simulator of that system are available.

In overview, the KO model consists of two components. First, any discrepancy between the physical system and the computer simulator is estimated; secondly, the most likely values of the calibration parameters are identified using prior information about these inputs and the data.

In detail, let  $\mathcal{X}$  denote the input space, and suppose  $n^p$  runs on a  $d$ -input physical system at control inputs  $\mathbf{x}_i^p = (x_{i,1}^p, \dots, x_{i,d}^p)^\top \in \mathcal{X}$ ,  $i = 1, \dots, n^p$ , yield observations denoted  $y^p(\mathbf{x}_1^p), \dots, y^p(\mathbf{x}_{n^p}^p)$ . Here  $^\top$  denotes transpose. Assume  $y^p(\mathbf{x}_i^p)$  can be described as a realization from

$$Y^p(\mathbf{x}_i^p) = \mu(\mathbf{x}_i^p) + \epsilon(\mathbf{x}_i^p), \quad (1)$$

where  $\mu(\mathbf{x}_i^p)$  denotes the mean response of the physical system at  $\mathbf{x}_i^p$  and  $\epsilon(\mathbf{x}_1^p), \dots, \epsilon(\mathbf{x}_{n^p}^p)$  are independent  $N(0, \sigma_\epsilon^2)$  measurement errors with unknown variance  $\sigma_\epsilon^2$  (or precision  $\lambda_\epsilon \equiv 1/\sigma_\epsilon^2$ ).

Let  $y^s(\mathbf{x}_i^s, \mathbf{t}_i)$ ,  $i = 1, \dots, n^s$ , denote the output from  $n^s$  runs of a deterministic computer simulator, where  $\mathbf{x}_i^s = (x_{i,1}^s, \dots, x_{i,d}^s)^\top \in \mathcal{X}$  is a  $d \times 1$  control input

vector, and  $\mathbf{t}_i = (t_{i,1}, \dots, t_{i,b})^\top$  is a  $b$ -dimensional calibration parameter required by the simulator. In this paper, the simulator data are modeled as realizations of the stationary GP:

$$Y^s(\mathbf{x}_i^s, \mathbf{t}_i) = \eta + Z(\mathbf{x}_i^s, \mathbf{t}_i), \quad i = 1, \dots, n^s, \quad (2)$$

where  $\eta$  is the (constant) mean function, and  $Z(\cdot, \cdot)$  is a zero-mean, stationary GP with variance  $\lambda_Z^{-1}$  and correlation function

$$R_Z((\mathbf{x}_i, \mathbf{t}_i), (\mathbf{x}_j, \mathbf{t}_j)) = \exp \left\{ - \sum_{k=1}^d \theta_k^x (x_{i,k} - x_{j,k})^2 - \sum_{e=1}^b \theta_e^t (t_{i,e} - t_{j,e})^2 \right\}; \quad (3)$$

$R_Z(\cdot, \cdot)$  is the Gaussian correlation function where  $(\theta_k^x, \theta_e^t)$  are positive parameters that control the smoothness of  $y^s(\mathbf{x}_i^s, \mathbf{t}_i)$  in  $(\mathbf{x}_i^s, \mathbf{t}_i)$ .

Following Kennedy and O'Hagan, define the bias (or "discrepancy") of the simulator at  $\mathbf{x}^p$  to be

$$\delta(\mathbf{x}^p) \equiv \mu(\mathbf{x}^p) - y^s(\mathbf{x}^p, \boldsymbol{\phi}). \quad (4)$$

Thus  $\delta(\mathbf{x}^p)$  measures the error in the simulator when run at  $\mathbf{t} = \boldsymbol{\phi}$  for the target mean  $\mu(\mathbf{x}^p)$ . In practice, simulators that are based on first principles regard  $\mu(\mathbf{x}_i^p)$  as  $\mu(\mathbf{x}_i^p, \boldsymbol{\phi})$  and see  $\boldsymbol{\phi}$  as a quantity in the physical system whose value can be solicited from subject-matter experts in the form of a prior distribution. In more rudimentary models,  $\mathbf{t}$  in  $y^s(\mathbf{x}^p, \mathbf{t})$  is better thought of as a tuning parameter and a "best value"  $\boldsymbol{\phi}$  minimizes the simulator error in some mathematical sense (see Han, Santner and Rawlinson (2009) or Tuo and Wu (2015) for examples of the latter).

The KO model describes the bias as realization of a zero mean GP,  $\Delta(\mathbf{x})$ , with covariance  $\lambda_\delta^{-1} \mathbf{R}_\delta$ , where the  $(i, j)^{th}$  element of  $\mathbf{R}_\delta$  is

$$R_\delta(\mathbf{x}_i, \mathbf{x}_j) = \exp \left\{ - \sum_{k=1}^d \theta_k^\delta (x_{i,k} - x_{j,k})^2 \right\}. \quad (5)$$

The discrepancy  $\Delta(\mathbf{x})$  is assumed to be stochastically independent of  $Z(\cdot, \cdot)$  and  $\epsilon(\cdot)$ .

As in Higdon et al. (2004), this paper places mutually independent priors on the unknown parameters  $\boldsymbol{\Omega} = (\boldsymbol{\lambda}, \boldsymbol{\theta}, \boldsymbol{\phi}, \eta)$  where  $\boldsymbol{\lambda}$  denotes the precision vector  $(\lambda_Z, \lambda_\delta, \lambda_\epsilon)^\top$ ,  $\boldsymbol{\theta}$  denotes the vector of combined smoothness parameters  $(\boldsymbol{\theta}^{x^\top}, \boldsymbol{\theta}^{t^\top}, \boldsymbol{\theta}^{\delta^\top})^\top$ , where  $\boldsymbol{\theta}^{x^\top} = (\theta_1^x, \dots, \theta_d^x)$ ,  $\boldsymbol{\theta}^{t^\top} = (\theta_1^t, \dots, \theta_b^t)$ , and

$\boldsymbol{\theta}^{\delta\top} = (\theta_1^\delta, \dots, \theta_d^\delta)$ ,  $\boldsymbol{\phi}$  denotes the unknown calibration parameter vector, and  $\eta$  the unknown mean of the GP for the simulator responses. In an abuse of notation, the  $\boldsymbol{\Omega}$  prior is denoted

$$\pi(\boldsymbol{\phi}) \times \pi(\boldsymbol{\theta}) \times \pi(\boldsymbol{\lambda}) \times \pi(\eta).$$

Bayarri, Berger, Paulo, Sacks, Cafeo, Cavendish et al. (2007) suggested using plug-in maximum likelihood estimators for the  $\boldsymbol{\theta}$  parameters in the GP model, arguing that this typically yields an answer similar to that of the fully Bayesian approach. However, this paper will use a fully Bayesian approach to prediction.

Let  $\mathbf{y}^p = (y^p(\mathbf{x}_1^p), \dots, y^p(\mathbf{x}_{n^p}^p))^\top$  denote the vector of the  $n^p$  physical observations and  $\mathbf{y}^s = (y^s(\mathbf{x}_1^s, \mathbf{t}_1), \dots, y^s(\mathbf{x}_{n^s}^s, \mathbf{t}_{n^s}))^\top$  the vector of the  $n^s$  computer simulator outputs. Combining physical observations and computer simulator outputs, let  $\mathcal{Y} = ((\mathbf{y}^p)^\top, (\mathbf{y}^s)^\top)^\top$ . Assuming (1)-(5), the mean of the physical system,  $\mu(\mathbf{x})$ , can be described as a realization of the sum

$$U(\mathbf{x}) \equiv Y^s(\mathbf{x}, \boldsymbol{\phi}) + \Delta(\mathbf{x}). \quad (6)$$

The posterior of  $\boldsymbol{\Omega}$  given the data  $\mathcal{Y}$  is formed in the usual way from the likelihood and the parameter prior. Let  $\mathbf{1}_{n^p+n^s}$  denote the  $(n^p + n^s) \times 1$  vector of 1s, and  $\mathbf{I}_{n^p}$  the  $n^p \times n^p$  identity matrix. The log-likelihood at  $(\boldsymbol{\lambda}, \boldsymbol{\theta}, \boldsymbol{\phi}, \eta)$  is the density

$$\ell(\boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \eta; \mathcal{Y}) \propto -\frac{1}{2} \log(|\Sigma_{\mathcal{Y}}|) - \frac{1}{2} \{(\mathcal{Y} - \eta \mathbf{1}_{n^p+n^s})^\top \Sigma_{\mathcal{Y}}^{-1} (\mathcal{Y} - \eta \mathbf{1}_{n^p+n^s})\}, \quad (7)$$

where

$$\Sigma_{\mathcal{Y}} = \lambda_Z^{-1} \begin{bmatrix} \mathbf{R}_Z^{pp} & \mathbf{R}_Z^{ps} \\ \mathbf{R}_Z^{sp} & \mathbf{R}_Z^{ss} \end{bmatrix} + \lambda_\delta^{-1} \begin{bmatrix} \mathbf{R}_\delta & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \lambda_\epsilon^{-1} \begin{bmatrix} \mathbf{I}_{n^p} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \quad (8)$$

Here  $\mathbf{R}_Z^{pp} = (R_Z((\mathbf{x}_i^p, \boldsymbol{\phi}), (\mathbf{x}_j^p, \boldsymbol{\phi})))$  is the  $n^p \times n^p$  correlation matrix corresponding to the physical system observations with  $R_Z(\cdot, \cdot)$  defined as in (3),  $\mathbf{R}_Z^{ps} = (R_Z((\mathbf{x}_i^p, \boldsymbol{\phi}), (\mathbf{x}_j^s, \mathbf{t}_j)))$  is the  $n^p \times n^s$  cross-correlation matrix,  $\mathbf{R}_Z^{sp} = (\mathbf{R}_Z^{ps})^\top$ ,  $\mathbf{R}_Z^{ss} = (R_Z((\mathbf{x}_i^s, \mathbf{t}_i), (\mathbf{x}_j^s, \mathbf{t}_j)))$  is the  $n^s \times n^s$  correlation matrix corresponding to the simulator data, and the components of  $\mathbf{R}_\delta$  are defined by (5). Thus, the posterior distribution of  $\boldsymbol{\Omega}$  satisfies

$$\pi(\boldsymbol{\Omega} \mid \mathcal{Y}) \propto \ell(\boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \eta; \mathcal{Y}) \times \pi(\boldsymbol{\phi}) \times \pi(\boldsymbol{\theta}) \times \pi(\boldsymbol{\lambda}) \times \pi(\eta). \quad (9)$$

Draws can be made from  $\pi(\boldsymbol{\Omega} \mid \mathcal{Y})$  using an adaptive MCMC based on the random-walk Metropolis-Hastings (M-H) algorithm; the adaptation adjusts the length of the random walk based on the acceptance rate in M-H (see Andrieu and Thoms (2008) for details regarding the adaptive MCMC).

To predict the mean response of the physical system at a new point  $\mathbf{x}_0$ , the following approximation is used

$$\begin{aligned}\hat{\mu}(\mathbf{x}_0) &= E\{U(\mathbf{x}_0) \mid \mathcal{Y}\} \\ &= E\{E\{U(\mathbf{x}_0) \mid \mathcal{Y}, \boldsymbol{\Omega}\}\} \\ &= \int E\{U(\mathbf{x}_0) \mid \mathcal{Y}, \boldsymbol{\Omega}\} \pi(\boldsymbol{\Omega} \mid \mathcal{Y}) d\phi d\boldsymbol{\theta} d\boldsymbol{\lambda} d\eta \\ &\approx \frac{1}{N_{mcmc}} \sum_{q=1}^{N_{mcmc}} E\{U(\mathbf{x}_0) \mid \mathcal{Y}, \boldsymbol{\Omega}^q\},\end{aligned}\quad (10)$$

where  $\boldsymbol{\Omega}^q = (\phi^q, \boldsymbol{\theta}^q, \boldsymbol{\lambda}^q, \eta^q)$  is the  $q^{th}$  draw from  $\pi(\boldsymbol{\Omega} \mid \mathcal{Y})$ .

The expectation in (10) is evaluated using the fact that the conditional distribution  $[U(\mathbf{x}_0) \mid \mathcal{Y}, \boldsymbol{\Omega}]$  is derived from the conditional distribution of the multivariate normal (see Supplementary Material) yielding

$$E\{U(\mathbf{x}_0) \mid \mathcal{Y}, \boldsymbol{\Omega}\} = \eta + \Sigma_{\mathbf{x}_0, \mathcal{Y}}^\top \Sigma_{\mathcal{Y}}^{-1} (\mathcal{Y} - \eta \mathbf{1}_{n^p+n^s}), \quad (11)$$

where  $\Sigma_{\mathcal{Y}}$  is defined in (8) and

$$\Sigma_{\mathbf{x}_0, \mathcal{Y}} = \lambda_Z^{-1} \begin{bmatrix} \mathbf{R}_Z^{0,p} \\ \mathbf{R}_Z^{0,s} \end{bmatrix} + \lambda_\delta^{-1} \begin{bmatrix} \mathbf{R}_\delta^{0,p} \\ \mathbf{0}_{n^s} \end{bmatrix}, \quad (12)$$

$\mathbf{R}_Z^{0,p}$  is the  $n^p \times 1$  vector of correlations having  $i^{th}$  component  $R_Z((\mathbf{x}_0, \boldsymbol{\phi}), (\mathbf{x}_i^p, \boldsymbol{\phi}))$ ,  $\mathbf{R}_Z^{0,s}$  is the  $n^s \times 1$  vector of correlations having  $i^{th}$  component  $R_Z((\mathbf{x}_0, \boldsymbol{\phi}), (\mathbf{x}_i^s, \mathbf{t}_i))$ ,  $\mathbf{R}_\delta^{0,p}$  is the  $n^p \times 1$  vector of correlations having  $i^{th}$  component  $R_\delta((\mathbf{x}_0, \boldsymbol{\phi}), (\mathbf{x}_i^p, \boldsymbol{\phi}))$ , and  $\mathbf{0}_{n^s}$  is the  $n^s \times 1$  vector of zeros, and

$$Var\{U(\mathbf{x}_0) \mid \mathcal{Y}, \boldsymbol{\Omega}\} = \lambda_Z^{-1} + \lambda_\delta^{-1} - \Sigma_{\mathbf{x}_0, \mathcal{Y}}^\top \Sigma_{\mathcal{Y}}^{-1} \Sigma_{\mathbf{x}_0, \mathcal{Y}}. \quad (13)$$

When a physical experiment and a computer simulator have multiple outputs, our modeling strategy is to assume that the  $(Z(\mathbf{x}^s, \mathbf{t}), \epsilon(\mathbf{x}^p), \Delta(\mathbf{x}^p))$  for the  $m$  output variables are mutually stochastically independent.

### 3 Pareto Minimization

Let  $w_\ell(\cdot)$ ,  $\ell = 1, \dots, m$ , denote  $m$  generic functions that are to be *minimized* and assume that all  $m$  outputs are equally important. Further, assume that the goal is to minimize all outputs simultaneously. This goal would be achievable if there were a single control vector (input point)  $\mathbf{x}^* \in \mathcal{X}$  satisfying  $w_\ell(\mathbf{x}^*) < w_\ell(\mathbf{x})$  for all  $\mathbf{x} \in \mathcal{X}$  and all outputs  $\ell = 1, \dots, m$ . In practice, however, a single solution to a multiobjective optimization problem need not exist. In that case, a set of compromise solutions is obtained via Pareto minimization.

**Definitions.** The point  $\mathbf{x}_i$  is said to *dominate* the point  $\mathbf{x}_j$ , written as  $\mathbf{x}_i \succ \mathbf{x}_j$ , if and only if (i)  $w_\ell(\mathbf{x}_i) \leq w_\ell(\mathbf{x}_j)$  for every  $\ell \in \{1, \dots, m\}$  and (ii) there exists at least one output  $\ell^* \in \{1, \dots, m\}$ , such that  $w_{\ell^*}(\mathbf{x}_i) < w_{\ell^*}(\mathbf{x}_j)$ . The vector  $\mathbf{x}^* \in \mathcal{X}$  is said to be a Pareto (optimal) minimizing point (or a nondominated point) provided there is no  $\mathbf{x} \in \mathcal{X}$  such that  $\mathbf{x}$  dominates  $\mathbf{x}^*$ . The Pareto Set (PS), denoted by  $\mathcal{P}_\mathcal{X}$ , is defined to be the collection of all Pareto minimizing points in  $\mathcal{X}$ ; the Pareto Front (PF), denoted by  $\mathcal{P}_\mathcal{Y}$ , is the image of  $\mathcal{P}_\mathcal{X}$  in the output space.

For simplicity, in the output space, we say that  $\mathbf{w}(\mathbf{x}_i)$  dominates  $\mathbf{w}(\mathbf{x}_j)$  if (i) and (ii) are satisfied. When seeking to minimize a set of functions  $w_\ell(\mathbf{x})$ ,  $\ell = 1, \dots, m$ , simultaneously, a point  $\mathbf{w}(\mathbf{x}_j)$  that is not on the PF cannot be the best choice, because it is dominated by at least one point on the PF.

As an example, consider a set of 15 points  $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{15}\}$  whose outputs  $\mathbf{w}(\mathbf{x}_i) = (w_1(\mathbf{x}_i), w_2(\mathbf{x}_i))$  are shown in Figure 1, where  $\mathbf{w}(\mathbf{x}_i)$  are depicted as filled diamonds clockwise for  $i = 1, \dots, 6$  and as open circles clockwise for  $i = 7, \dots, 15$ . For any point  $\mathbf{x}_j$ ,  $7 \leq j \leq 15$ , there is at least one  $\mathbf{x}_i$ ,  $1 \leq i \leq 6$ , that yields a smaller value of  $w_1(\cdot)$  and  $w_2(\cdot)$ , so every input vector yielding a circle point is dominated by at least one input vector yielding a diamond point. On the other hand, no two input points  $\mathbf{x}_{i_1}, \mathbf{x}_{i_2}$ ,  $1 \leq i_1 \neq i_2 \leq 6$ , dominate each other and these six input points form the PS,  $\mathcal{P}_\mathcal{X}$ , among the 15 input points; the corresponding six output points  $\mathbf{w}(\mathbf{x}_i)$ ,  $i = 1, \dots, 6$  form the PF,  $\mathcal{P}_\mathcal{Y}$ .

When the set of points in  $\mathcal{X}$  is too large to be evaluated by the simulator, an *approximate (estimated)* PF is taken to be the set of non-dominated points from among those output points predicted at the input points of a given design. The

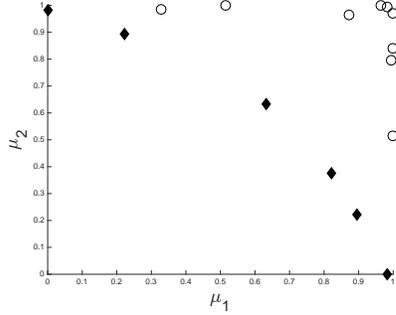


Figure 1: Diamonds are points on the PF in the set of points denoted by diamonds or circles. Each circle is dominated by one or more points on the PF.

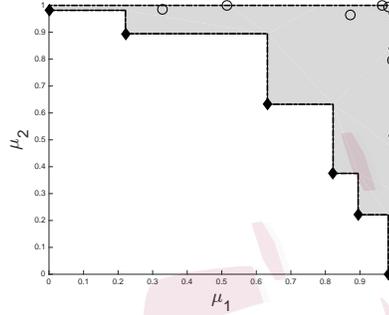


Figure 2: The area represents the hypervolume indicator for the PF of Figure 1, and the reference point  $\mathbf{r}$  is the filled square.

set of input points  $\mathbf{x}$  for which  $\mathbf{w}(\mathbf{x})$  is on the approximate PF is then called the *approximate, or current, PS*. An important issue in multiobjective optimization is the evaluation of the quality of an approximate (estimated) PF. In the sequential procedure proposed in Section 4, the approximate PF is updated after each subsequent output value is predicted. Although the true PF is unknown, the improvement (deterioration) of the approximation can be assessed via the increase (decrease) in the *hypervolume indicator (HI)*. The HI indicator was introduced by Zitzler and Thiele (1999) and is a popular real-valued quality measure used to assess PF accuracy (see, also, Zitzler, Knowles and Thiele, 2008); it is defined as follows.

First, let  $w_\ell(\mathbf{x})$  be the output of the  $\ell^{th}$  function and  $\mathcal{X}$  the common domain for the problem, and select a “reference point”  $\mathbf{r}$  with  $\ell^{th}$  coordinate satisfying  $r_\ell \geq \sup_{\mathcal{X}} w_\ell(\mathbf{x})$ . However, when  $\mathbf{w}(\mathbf{x})$  has been evaluated for a finite set of inputs, the reference point requirement is often relaxed by setting  $r_\ell$  equal to maximum of the  $w_\ell(\mathbf{x})$  over the inputs. Then, intuitively, the HI of an approximate PF,  $\mathcal{P}_y$ , is the area or volume of the set of points in objective space that both dominate  $\mathbf{r}$  and are dominated by one or more points of  $\mathcal{P}_y$ . The shaded area in Figure 2 is the HI for the (approximate) PF of Figure 1.

## 4 Proposed Sequential Design Methodology

This section proposes a sequential design methodology for identifying Pareto minimizing solutions for the mean of a physical system having multiple outputs using a calibrated simulator. Section 4.1 defines a maximin fitness function which will be used for selecting new settings of the control input variables.

### 4.1 Minimax Fitness Function

Jones, Schonlau and Welch (1998) and Ranjan, Bingham and Michailidis (2008) give early proposals of sequential designs based on “improvement functions”; these papers guide the search for the maximum of a simulator surface and a specified contour of it, respectively. In settings with multiple simulator outputs Keane (2006) and Svenson and Santner (2016) proposed improvement criteria to guide the sequential choice of inputs for determining Pareto minimizing solutions. Specifically, the Svenson and Santner (2016) improvement criterion is based on a “fitness function.” Here, we propose a modification of fitness-based improvement methodology for finding Pareto Fronts based on predicted mean values from a statistically calibrated simulator.

To describe the methodology of this paper, initially assume that the mean function  $\boldsymbol{\mu}(\boldsymbol{x}) = (\mu_1(\boldsymbol{x}), \dots, \mu_m(\boldsymbol{x}))$  of an  $m$ -output physical system is *known*. Define  $\mathcal{P}_X^{n^p+n^s}$  to be the current PS based on the calibrated simulator predictions at the union of the control inputs for the first  $n^p$  physical system observations and the control portions of the first  $n^s$  inputs for the simulation runs. The *minimax fitness function* of  $\boldsymbol{\mu}(\cdot)$  at  $\boldsymbol{x} \in \mathcal{X}$  is then defined to be

$$I_{\mathcal{F}}(\boldsymbol{\mu}(\boldsymbol{x})) \equiv \left[ \min_{\boldsymbol{x}_i \in \mathcal{P}_X^{n^p+n^s}} \max_{\ell=1, \dots, m} (\mu_{\ell}(\boldsymbol{x}_i) - \mu_{\ell}(\boldsymbol{x})) \right]_+, \quad (14)$$

where  $a_+$  denotes the positive part of  $a$ , i.e.,  $a_+ = \max\{a, 0\}$ . The intuition for (14) as follows. The difference  $\mu_{\ell}(\boldsymbol{x}_i) - \mu_{\ell}(\boldsymbol{x})$  is the improvement (or degradation) in the  $\ell^{th}$  output at the new point,  $\boldsymbol{x}$ , compared with that at the  $i^{th}$  point,  $\boldsymbol{x}_i$ , in the current PS. The maximum in (14) identifies the output  $\mu_{\ell}(\cdot)$ ,  $\ell = 1, \dots, m$ , producing maximum improvement at  $\boldsymbol{x}$  relative to  $\mu_{\ell}(\boldsymbol{x}_i)$ . If the minimum in (14) over the points  $\boldsymbol{x}_i$  in the current PS is *positive*, then  $\boldsymbol{x}$  is not dominated by any point in the current PS.

The greater  $I_{\mathcal{F}}(\boldsymbol{\mu}(\boldsymbol{x}))$ , the more desirable is  $\boldsymbol{x}$ . Thus an intuitive choice of additional  $\boldsymbol{x} \in \mathcal{X}$  to add to a current set of training inputs is that  $\boldsymbol{x}$  which maximizes  $I_{\mathcal{F}}(\boldsymbol{\mu}(\boldsymbol{x}))$ . Note that, since (14) is based on only those inputs in the current PS,  $\mathcal{P}_X^{n^p+n^s}$ , rather than all  $n^p + n^s$  design points, the computational expense is reduced.

In this paper,  $\boldsymbol{\mu}(\cdot)$  is unknown and is predicted using  $\boldsymbol{U}(\cdot)$  in (6). The improvement at a new  $\boldsymbol{x}$  is assessed by the conditional expected minimax fitness function given the current data  $\mathcal{Y}^c$ ; that is

$$E\{I_{\mathcal{F}}(\boldsymbol{U}(\boldsymbol{x})) \mid \mathcal{Y}^c\}, \quad (15)$$

where  $\mathcal{Y}^c = (\mathcal{Y}_1^\top, \dots, \mathcal{Y}_m^\top)^\top$ , and  $\mathcal{Y}_\ell = ((\boldsymbol{y}_\ell^p)^\top, (\boldsymbol{y}_\ell^s)^\top)^\top$  for  $\ell = 1, \dots, m$ . In the Supplementary Material it is shown that when the mean function  $\boldsymbol{\mu}(\boldsymbol{x})$  is known, the minimax fitness function (14) is zero for any  $\boldsymbol{x}$  where the simulator has been previously run or a physical experiment conducted. When  $\boldsymbol{\mu}(\boldsymbol{x})$  is predicted as in (10) using the expected  $\boldsymbol{U}(\boldsymbol{x})$ , then the expected minimax fitness function (15) need not be zero due to the presence of measurement error or bias in the output, depending on the experimental platform used. However, in our examples, maximization of (15) has not led to exact duplication of previously selected inputs.

Section 4.2 gives an algorithm for constructing sequential designs for identifying the PF in two cases: (i) when additional observations can only be taken on the *physical system*, and (ii) when additional runs can only be made from the *simulator of the physical system*. The case where the researcher has the flexibility to choose whether to run the simulator or to observe the physical system at any stage is discussed briefly in Section 6.

## 4.2 Pareto Minimization Algorithms

Consider the setting in which a researcher can only make additional observations on the physical system and the set of simulator runs is fixed. Assume that *initial training data* is available in the form of observations from the physical system and runs of the simulator based on the initial designs  $\boldsymbol{X}_{n^p}$  and  $\boldsymbol{X}_{n^s}$ , respectively. In the examples of Section 5, the initial training designs are maximin Latin Hypercube Designs.

**Step 1.1** Fit the Bayesian calibrated model (6) for each output independently and predict the vector of means of the physical system observations at  $\mathbf{X}_{n^p}$  and  $\mathbf{X}_{n^s}$ .

**Step 1.2** Determine the PS,  $\mathcal{P}_X^{n^p+n^s}$ , based on the predictions in Step 1.1.

**Step 1.3** Select the next vector  $\mathbf{x}^*$  of control input settings to maximize the posterior expected minimax fitness function given the current data  $\mathcal{Y}^c$ , i.e.,

$$\mathbf{x}^* \in \operatorname{argmax}_{\mathbf{x} \in \mathcal{X}} E\{I_{\mathcal{F}}(\mathbf{U}(\mathbf{x})) \mid \mathcal{Y}^c\}. \quad (16)$$

**Step 1.4** Observe  $\mathbf{y}^p(\mathbf{x}^*)$  and form the augmented matrix of physical system data  $\mathbf{y}_{n^p+1}^p = ((\mathbf{y}_{n^p}^p)^\top, (\mathbf{y}^p(\mathbf{x}^*))^\top)^\top$ . The details for the calculation of (16) are given in Section 4.3.

**Step 1.5** Set  $n^p = n^p + 1$ , and go to Step 1.1 unless the sampling budget has been exhausted or a stopping criterion has been met.

Now consider the case when a researcher can only make additional runs of the simulator while no further observations can be made using the physical system. The proposed sequential design uses Steps 1.1–1.3 to select a control input  $\mathbf{x}^*$ . To run the simulator, a calibration parameter denoted  $\mathbf{t}^*$ , must also be determined; Steps 2.4 and 2.5 are used to select  $\mathbf{t}^*$ .

**Steps 2.1–2.3** Perform Steps 1.1–1.3 to select  $\mathbf{x}^*$

**Step 2.4** Construct  $\mathbf{t}^*$  to minimize the sum of the posterior mean squared prediction errors (MSPE) given the current data  $\mathcal{Y}^c$  and the selected  $\mathbf{x}^*$ , i.e.

$$\mathbf{t}^* \in \operatorname{argmin}_{\mathbf{t} \in \mathcal{T}} \sum_{\ell=1}^m E\{[\hat{U}_\ell(\mathbf{x}^*) - U_\ell(\mathbf{x}^*)]^2 \mid \mathcal{Y}^c\}, \quad (17)$$

where  $\hat{U}_\ell(\mathbf{x}^*)$  is the posterior mean of  $[U_\ell(\mathbf{x}^*) \mid \mathcal{Y}_\ell, Y_\ell^s(\mathbf{x}^*, \mathbf{t})]$ . The mean,  $\hat{U}_\ell(\mathbf{x}^*)$ , is a function of  $\mathbf{t}$  through  $Y_\ell^s(\mathbf{x}^*, \mathbf{t})$ ;  $\hat{U}_\ell(\mathbf{x}^*)$  can be expressed in closed form (see Section 4.4 for additional details).

**Step 2.5** Evaluate  $\mathbf{y}^s(\mathbf{x}^*, \mathbf{t}^*)$  and form the augmented matrix of simulator data  $\mathbf{y}_{n^s+1}^s = ((\mathbf{y}_{n^s}^s)^\top, (\mathbf{y}^s(\mathbf{x}^*, \mathbf{t}^*))^\top)^\top$ .

**Step 2.6** Set  $n^s = n^s + 1$ , and go to Step 2.1 unless the sampling budget has been exhausted or a stopping criterion has been met.

### 4.3 Calculation of the Expected Minimax Fitness Function

Applying the Bayesian framework, the iterated expectation formula and law of large numbers show that the following approximation holds

$$E\{I_{\mathcal{F}}(\mathbf{U}(\mathbf{x})) \mid \mathcal{Y}^c\} \approx \frac{1}{N_{mcmc}} \sum_{q=1}^{N_{mcmc}} E\{I_{\mathcal{F}}(\mathbf{U}(\mathbf{x})) \mid \mathcal{Y}^c, \boldsymbol{\Omega}^q\}, \quad (18)$$

where  $\boldsymbol{\Omega}^q$ ,  $1 \leq q \leq N_{mcmc}$ , are draws from the posterior distribution  $[\boldsymbol{\Omega} \mid \mathcal{Y}^c]$ , and  $I_{\mathcal{F}}(\mathbf{U}(\mathbf{x}))$  is defined in Section 4.1. The evaluation of  $E\{I_{\mathcal{F}}(\mathbf{U}(\mathbf{x})) \mid \mathcal{Y}^c, \boldsymbol{\Omega}\}$  can be provided in near-closed form as follows. Calculation gives

$$E\{I_{\mathcal{F}}(\mathbf{U}(\mathbf{x})) \mid \mathcal{Y}^c, \boldsymbol{\Omega}\} = E \left\{ \max \left( 0, \min_{\mathbf{x}_i \in \mathcal{P}_X^{n^p+n^s}} \max_{1 \leq \ell \leq m} U_{\ell}^*(\mathbf{x}_i, \mathbf{x}) \right) \mid \mathcal{Y}^c, \boldsymbol{\Omega} \right\} \quad (19)$$

$$= E \{ \max(0, U_{\min}^*(\mathbf{x})) \mid \mathcal{Y}^c, \boldsymbol{\Omega} \} \quad (20)$$

where  $U_{\ell}^*(\mathbf{x}_i, \mathbf{x}) = U_{\ell}(\mathbf{x}_i) - U_{\ell}(\mathbf{x})$  and

$$U_{\min}^*(\mathbf{x}) = \min_{\mathbf{x}_i \in \mathcal{P}_X^{n^p+n^s}} \max_{1 \leq \ell \leq m} U_{\ell}^*(\mathbf{x}_i, \mathbf{x}).$$

Because conditional on  $(\mathcal{Y}^c, \boldsymbol{\Omega})$ ,  $U_{\ell}(\mathbf{x}_i)$  and  $U_{\ell}(\mathbf{x})$  are bivariate normally distributed random variables,  $U_{\ell}^*(\mathbf{x}_i, \mathbf{x})$  is normally distributed with mean

$$\mu_{\ell}^*(\mathbf{x}_i, \mathbf{x}) = (\Sigma_{\ell, \mathbf{x}_i, \mathcal{Y}} - \Sigma_{\ell, \mathbf{x}, \mathcal{Y}})^{\top} \Sigma_{\ell, \mathcal{Y}}^{-1} (\mathcal{Y}_{\ell} - \eta_{\ell} \mathbf{1}_{n^p+n^s}) \quad (21)$$

and variance

$$(\sigma_{\ell}^*(\mathbf{x}_i, \mathbf{x}))^2 = 2(\lambda_{\ell, Z}^{-1} + \lambda_{\ell, \delta}^{-1}) - \Sigma_{\ell, \mathbf{x}_i, \mathcal{Y}}^{\top} \Sigma_{\ell, \mathcal{Y}}^{-1} \Sigma_{\ell, \mathbf{x}_i, \mathcal{Y}} - \Sigma_{\ell, \mathbf{x}, \mathcal{Y}}^{\top} \Sigma_{\ell, \mathcal{Y}}^{-1} \Sigma_{\ell, \mathbf{x}, \mathcal{Y}} - 2(\Sigma_{\ell, \mathbf{x}_i, \mathbf{x}} - \Sigma_{\ell, \mathbf{x}_i, \mathcal{Y}}^{\top} \Sigma_{\ell, \mathcal{Y}}^{-1} \Sigma_{\ell, \mathbf{x}, \mathcal{Y}}), \quad (22)$$

where

$$\Sigma_{\ell, \mathbf{x}_i, \mathbf{x}} = \lambda_{\ell, Z}^{-1} R_{\ell, Z}((\mathbf{x}_i, \boldsymbol{\phi}_{\ell}), (\mathbf{x}, \boldsymbol{\phi}_{\ell})) + \lambda_{\ell, \delta}^{-1} R_{\ell, \delta}((\mathbf{x}_i, \boldsymbol{\phi}_{\ell}), (\mathbf{x}, \boldsymbol{\phi}_{\ell})), \quad (23)$$

(see Supplementary Material), and the notation  $R_{\ell, Z}$ ,  $R_{\ell, \delta}$ ,  $\Sigma_{\ell, \mathcal{Y}}$  and  $\Sigma_{\ell, \mathbf{x}, \mathcal{Y}}$  denotes (3), (5), (8) and (13), respectively, for the  $\ell^{th}$  output.

The distribution of  $\max_{1 \leq \ell \leq m} U_{\ell}^*(\mathbf{x}_i, \mathbf{x})$  in (19) can be derived as follows. In general the distribution of the maximum of several normal random variables

need not be normally distributed. Let  $\phi(\cdot)$  and  $\Phi(\cdot)$  denote the probability density function (PDF) and the cumulative distribution function (CDF) of the standard normal distribution, respectively. Since independent Bayesian calibration models described in Section 2 are fit to each output, the PDFs and CDFs of  $\max_{1 \leq \ell \leq m} U_\ell^*(\mathbf{x}_i, \mathbf{x})$  can be calculated using Lemma 1.

**Lemma 1.** *Suppose that  $V_1, \dots, V_m$  are mutually independent Gaussian random variables with mean  $\mu_\ell$  and standard deviation  $\sigma_\ell$ , for  $\ell = 1, \dots, m$ , respectively. Let  $V = \max_{\ell=1, \dots, m}(V_\ell)$ . Then the PDF of  $V$  is*

$$f_V(v) = \sum_{\ell=1}^m \left[ \frac{1}{\sigma_\ell} \phi\left(\frac{v - \mu_\ell}{\sigma_\ell}\right) \prod_{j=1, j \neq \ell}^m \Phi\left(\frac{v - \mu_j}{\sigma_j}\right) \right], \quad (24)$$

and the CDF of  $V$  is

$$F_V(v) = \prod_{\ell=1}^m \Phi\left(\frac{v - \mu_\ell}{\sigma_\ell}\right). \quad (25)$$

Finally,  $E\{I_{\mathcal{F}}(\mathbf{U}(\mathbf{x})) \mid \mathcal{Y}^c, \boldsymbol{\Omega}\}$  can be evaluated once the PDF of  $U_{\min}^*(\mathbf{x})$  in (20) is determined; this density can be calculated using Lemma 2.

**Lemma 2.** *Suppose that  $W_1, \dots, W_m$  are mutually independent random variables with PDF,  $f_\ell(w_\ell)$ , and CDF,  $F_\ell(w_\ell)$ , for  $\ell = 1, \dots, m$ , respectively. Let  $W = \min_{\ell=1, \dots, m}(W_\ell)$ . Then, the PDF of  $W$  is*

$$f_W(w) = \sum_{\ell=1}^m \left[ f_\ell(w) \prod_{j=1, j \neq \ell}^m (1 - F_j(w)) \right]. \quad (26)$$

Our implementation uses (26) to evaluate  $E\{I_{\mathcal{F}}(\mathbf{U}(\mathbf{x})) \mid \mathcal{Y}^c, \boldsymbol{\Omega}\}$  numerically using standard integration techniques implemented in the MATLAB function `integral.m`.

#### 4.4 The Calculation of the MSPE

Given the update  $\mathbf{x}^*$  for the control portion of the simulator input, the selection of the update of the calibration parameter requires the evaluation of the sum of the expectations in (17). Recall that in (17),  $\hat{U}_\ell(\mathbf{x}^*)$  denotes the posterior mean of  $[U_\ell(\mathbf{x}^*) \mid \mathcal{Y}^c, Y_\ell^s(\mathbf{x}^*, \mathbf{t})]$ . Let  $Q_\ell(\mathbf{x}^*) = [\hat{U}_\ell(\mathbf{x}^*) - U_\ell(\mathbf{x}^*)]^2$  be the

squared prediction error for the  $\ell^{th}$  output at  $\mathbf{x}^*$ . The expectation in (17) can be approximated by

$$E\{Q_\ell(\mathbf{x}^*)|\mathcal{Y}^c\} \approx \frac{1}{N_{mcmc}} \sum_{q=1}^{N_{mcmc}} E\{Q_\ell(\mathbf{x}^*) | \mathcal{Y}^c, \mathbf{\Omega}^q\}. \quad (27)$$

Conditioning on the output at a potential new  $\mathbf{t}$ , the expectation in (27) can be written

$$E\{Q_\ell(\mathbf{x}^*) | \mathcal{Y}^c, \mathbf{\Omega}\} = E_{[Y_\ell^s(\mathbf{x}^*, \mathbf{t}) | \mathcal{Y}^c, \mathbf{\Omega}]} E\{Q_\ell(\mathbf{x}^*) | \mathcal{Y}^c, Y_\ell^s(\mathbf{x}^*, \mathbf{t}), \mathbf{\Omega}\}. \quad (28)$$

This method of conditioning is termed ‘‘Stepwise uncertainty reduction’’ and has been used by numerous authors (see, for example, Williams, Santner and Notz, 2000; Picheny, 2015). The inner expectation in (28) depends on the data in  $\mathcal{Y}^c$  only through the  $\ell^{th}$  output which is  $\mathcal{Y}_\ell$ . Thus

$$E\{Q_\ell(\mathbf{x}^*) | \mathcal{Y}^c, Y_\ell^s(\mathbf{x}^*, \mathbf{t}), \mathbf{\Omega}\} = Var\{U_\ell(\mathbf{x}^*)|\mathcal{Y}_\ell, Y_\ell^s(\mathbf{x}^*, \mathbf{t}), \mathbf{\Omega}\}.$$

This variance is

$$Var\{U_\ell(\mathbf{x}^*)|\mathcal{Y}_\ell, Y_\ell^s(\mathbf{x}^*, \mathbf{t}), \mathbf{\Omega}\} = \lambda_{\ell, Z}^{-1} + \lambda_{\ell, \delta}^{-1} - \Sigma_{\ell, \mathbf{x}^*, \mathcal{Y}, Y_\ell^s(\mathbf{x}^*, \mathbf{t})}^{-1} \Sigma_{\ell, \mathcal{Y}, Y_\ell^s(\mathbf{x}^*, \mathbf{t})}^{-1} \Sigma_{\ell, \mathbf{x}^*, \mathcal{Y}, Y_\ell^s(\mathbf{x}^*, \mathbf{t})}, \quad (29)$$

from (13), where the conditioning data are  $[\mathcal{Y}_\ell, Y_\ell^s(\mathbf{x}^*, \mathbf{t})]$ , and where

$$\begin{aligned} \Sigma_{\ell, \mathbf{x}^*, \mathcal{Y}, Y_\ell^s(\mathbf{x}^*, \mathbf{t})} &= \left[ \Sigma_{\ell, \mathbf{x}^*, \mathcal{Y}} \quad \lambda_{\ell, Z}^{-1} R_{\ell, Z}((\mathbf{x}^*, \phi_\ell), (\mathbf{x}^*, \mathbf{t})) \right], \quad \text{and} \\ \Sigma_{\ell, \mathcal{Y}, Y_\ell^s(\mathbf{x}^*, \mathbf{t})} &= \lambda_{\ell, Z}^{-1} \begin{bmatrix} \mathbf{R}_{\ell, Z}^{pp} & \mathbf{R}_{\ell, Z}^{ps} & \mathbf{R}_{\ell, Z}^{pt} \\ \mathbf{R}_{\ell, Z}^{sp} & \mathbf{R}_{\ell, Z}^{ss} & \mathbf{R}_{\ell, Z}^{st} \\ \mathbf{R}_{\ell, Z}^{tp} & \mathbf{R}_{\ell, Z}^{ts} & 1 \end{bmatrix} + \lambda_{\ell, \delta}^{-1} \begin{bmatrix} \mathbf{R}_{\ell, \delta} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ &+ \lambda_{\ell, \epsilon}^{-1} \begin{bmatrix} \mathbf{I}_{n^p} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \end{aligned} \quad (30)$$

The definitions of  $\mathbf{R}_{\ell, Z}^{pp}$ ,  $\mathbf{R}_{\ell, Z}^{ps}$ ,  $\mathbf{R}_{\ell, Z}^{ss}$ ,  $\mathbf{R}_{\ell, Z}^{sp}$ , and  $\mathbf{R}_{\ell, \delta}$  can be found in (8) for  $\ell^{th}$  output. The symbol  $\mathbf{R}_{\ell, Z}^{pt}$  is the  $n^p \times 1$  vector of correlations ( $R_{\ell, Z}((\mathbf{x}_i^p, \phi_\ell), (\mathbf{x}^*, \mathbf{t}))$ ), and  $\mathbf{R}_{\ell, Z}^{st}$  is the  $n^s \times 1$  vector of correlations ( $R_{\ell, Z}((\mathbf{x}_i^s, \mathbf{t}_i), (\mathbf{x}^*, \mathbf{t}))$ ). In addition,  $\mathbf{R}_{\ell, Z}^{sp} = (\mathbf{R}_{\ell, Z}^{ps})^\top$ ,  $\mathbf{R}_{\ell, Z}^{tp} = (\mathbf{R}_{\ell, Z}^{pt})^\top$ , and  $\mathbf{R}_{\ell, Z}^{ts} = (\mathbf{R}_{\ell, Z}^{st})^\top$ .

Because the inner expectation in (28), which is calculated by (29), is independent of  $Y_\ell^s(\mathbf{x}^*, \mathbf{t})$  (but depends on the inputs  $\mathbf{x}^*$  and  $\mathbf{t}$ ), the expectation in (27) can be calculated as

$$\frac{1}{N_{mcmc}} \sum_{q=1}^{N_{mcmc}} \text{Var}\{U_\ell(\mathbf{x}^*) | \mathcal{Y}_\ell, Y_\ell^s(\mathbf{x}^*, \mathbf{t}), \boldsymbol{\Omega}^q\}. \quad (31)$$

## 5 Examples

This section demonstrates the performance of the proposed sequential design methodology for two analytic test functions that are adapted from the multiple-objective optimization literature. In each case, the mean of the physical system is taken to be the multiple-objective function for the given system while the bias (and hence simulator output) is constructed separately. The first example has  $d = 2$  control inputs,  $b = 1$  calibration input, and  $m = 2$  outputs; the second example has  $d = 4$  control inputs,  $b = 1$  calibration input, and  $m = 4$  outputs.

For both examples, the initial designs for the physical system and simulator experiments were taken to be maximin LHDs (obtained from the website <https://spacefillingdesigns.nl>). The expectation in (11) was evaluated using MATLAB code written by the first author. The MATLAB function `paretoset.m` was used to calculate  $\mathcal{P}_X^{n^p+n^s}$  in (14) (Cao, 2009). The optimizations in (16) and (17) were performed using a hybrid algorithm that identified a starting point for a quasi-Newton search by using the Dividing Rectangles DIRECT global search algorithm of (Finkel, 2003) and the optimizer `fmincon.m` from the MATLAB Optimization toolbox.

All calculations in (11) were made using  $N_{mcmc} = 250$  equally-spaced samples drawn from MCMC runs with 5,000 burn-in and 5,000 production draws from the posterior distribution of  $\boldsymbol{\Omega}$ . As in Higdon, Kennedy, Cavendish, Cafeo and Ryne (2004), the software used to fit the model of this paper assumes two standardizations that help provide numerical stability in the calculations and facilitate the choice of priors. First, each input was standardized to  $[0, 1]$ . Secondly, the mean and standard deviations of the  $m$  simulator outputs were used to standardize both the simulator outputs and the corresponding physical system outputs so that the former had sample mean 0 and sample variance 1. Thus the

prior for  $Y^s(\mathbf{x}, t)$  was taken to be a *zero mean* GP with unknown precision  $\lambda_Z$  that was modeled with prior mean 1. The calibration parameter  $\phi$  was assumed to have the (non-informative) normal prior with mean .5 and standard deviation .25. All  $\theta$  parameters for  $Y^s(\mathbf{x}, t)$  and  $\Delta(\mathbf{x})$  were given priors induced by assuming the components of  $\exp\{-4\theta\}$  are independent and identically Beta(5, 5) distributed. Let  $\Gamma(a, b)$  denote the Gamma distribution with mean  $a/b$  and variance  $a/b^2$ . The prior distributions of the  $\lambda$  components were assumed to be independent with  $\pi(\lambda_Z) \sim \Gamma(5, 5)$ ,  $\pi(\lambda_\delta) \sim \Gamma(12, 2)$ , and  $\pi(\lambda_\epsilon) \sim \Gamma(40, 4)$ .

### 5.1 The *MOP2* Function

Introduced by Fonseca and Fleming (1995), the *MOP2* function has  $m = 2$  outputs and  $d = 2$  inputs. The input space is  $(x_1, x_2) \in [-2, 2]^2$  and the outputs, used as the mean of the physical system, are

$$\mu_\ell^p(\mathbf{x}) = 1 - \exp\left\{-\sum_{k=1}^2 \left(x_k + \frac{(-1)^\ell}{\sqrt{2}}\right)^2\right\}, \quad \text{for } \ell = 1, 2. \quad (32)$$

The goal was to identify the input settings that minimize  $\mu_1^p(\mathbf{x})$  and  $\mu_2^p(\mathbf{x})$  in the Pareto sense.

The physical system observations were generated from  $Y_\ell^p(\mathbf{x}) = \mu_\ell^p(\mathbf{x}) + \epsilon_\ell$ ,  $\ell = 1, 2$ , where  $\epsilon_\ell \sim N(0, 0.05^2)$ . The simulator runs were taken to have a single calibration input,  $t \in [-2, 2]$ . The simulator output was  $y_\ell^s(\mathbf{x}, t) = \mu_\ell^p(\mathbf{x}) + |t|/50 + .4 + (x_1x_2/20)^2$  for  $\ell = 1, 2$ . The parameter  $\phi$  was taken to be 0 so that the bias function was  $\delta(\mathbf{x}) = \mu_\ell^p(\mathbf{x}) - y_\ell^s(\mathbf{x}, 0) = -.4 - (x_1x_2/20)^2$ . Thus the simulator output  $y_\ell^s(\mathbf{x}, 0)$  was biased for the target  $\mu_\ell^p(\mathbf{x})$ .

The true PS of  $\mu_1^p(\mathbf{x})$  and  $\mu_2^p(\mathbf{x})$  consists of all  $\mathbf{x}$  in

$$\mathcal{P}_X = \left\{ \mathbf{x}_k : x_1 = x_2 \text{ and } -\frac{1}{\sqrt{2}} \leq x_1, x_2 \leq \frac{1}{\sqrt{2}} \right\}. \quad (33)$$

In the assessment of the performance of our sequential method,  $\mathcal{P}_X$  is approximated by 100  $\mathbf{x}$  points uniformly spread over the line segment in (33) and the PF is approximated by the images of these points.

For both physical and computer experiments, the size of the initial maximin LHDs was chosen to be  $5 \times$  the number of input dimensions, i.e. an initial 10 points for the physical experiment and 15 points for the computer simulator.

These initial designs were augmented sequentially using the algorithm in Section 4 with *10 new inputs* either by adding physical observations (Case 1) or simulator runs (Case 2).

The left and right panels of Figure 3 plot  $(\mu_1^p(\mathbf{x}), \mu_2^p(\mathbf{x}))$  for 10 added points from the proposed method for Cases 1 and 2, respectively. For either sampling option, simulator or physical system, the  $(\mu_1^p(\mathbf{x}), \mu_2^p(\mathbf{x}))$  at each  $\mathbf{x}$  identified by the algorithms are close to the true PF and spread over the entire PF. Thus the method provides many good options for the experimenter.

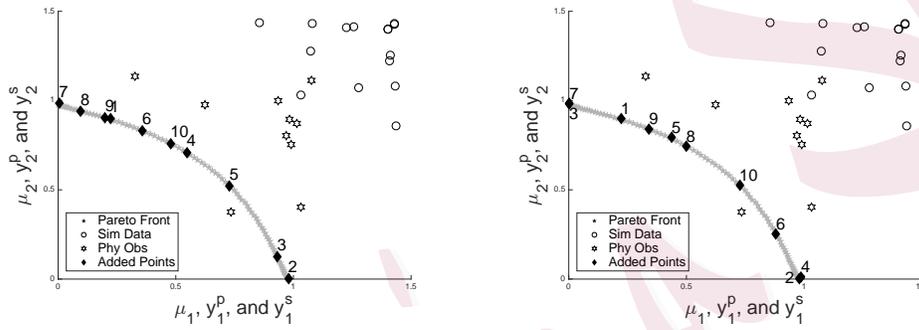


Figure 3: Left Panel:  $(\mu_1^p(\mathbf{x}), \mu_2^p(\mathbf{x}))$  for the 10 points identified by the Sequential Pareto minimization procedure that adds only physical observations are labeled 1, . . . 10 and denoted by filled diamonds. Right Panel:  $(\mu_1^p(\mathbf{x}), \mu_2^p(\mathbf{x}))$  for the 10 points identified by the Sequential Pareto minimization procedure that adds only simulator runs are labeled 1, . . . 10 and denoted by filled diamonds. In both panels, the initial observations of physical experiment are denoted by stars (\*), and the initial runs of computer simulator are denoted by circles (o)

In practice, the mean functions of the physical systems are unknown and so the quality of the PS and PF identified by the proposed procedure was assessed by the following method. After each output was added, using either the Case 1 or the Case 2 procedure, the simulator was used to predict  $(\mu_1^p(\mathbf{x}), \mu_2^p(\mathbf{x}))$  at the control portion of the inputs for all initial data and all data added up to that stage. Then, from among these predicted physical means, the set of non-dominated predicted means was identified, together with their corresponding inputs; these inputs were treated as the approximate PS. The hypervolume indicator (HI) was calculated relative to the reference point  $\mathbf{r} = (1, 1)$  for the set of  $(\mu_1^p(\mathbf{x}), \mu_2^p(\mathbf{x}))$  evaluated at the approximate Pareto Set. The left panel of Figure 4 shows the HI for the resulting set of points for the proposed method for the Case 1 and

Case 2 sampling plans. It also plots the maximum HI which is obtained from the true PF. As new data is acquired the HI of the non-dominated points approaches the HI of the true PF. This figure, and other examples not shown, suggests that adding physical system observations can be more valuable than adding simulator runs.

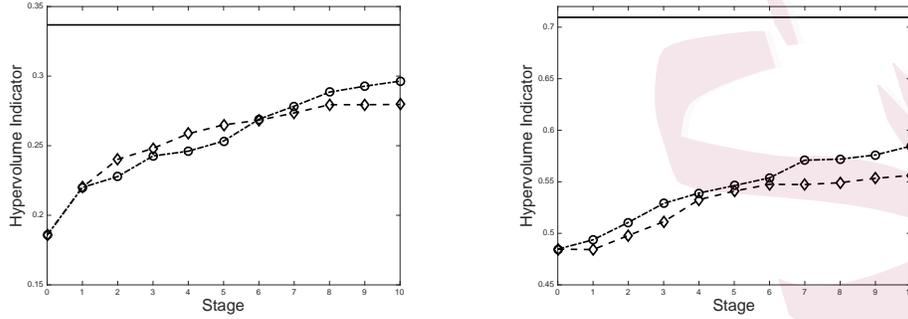


Figure 4: In both panels: the top solid horizontal line is the hypervolume indicator of the true Pareto Front. The two line segments show the hypervolume indicator of the set of  $(\mu_1^p(\mathbf{x}), \mu_2^p(\mathbf{x}))$  evaluated at the approximate Pareto Set at each stage, 1 to 10, of the Sequential Pareto minimization procedure when adding only physical observations (circles) or when adding only simulator runs (diamonds). Left Panel: the *MOP2* example; Right Panel: the *DTLZ2* example.

## 5.2 The *DTLZ2* Function

This example evaluates the performance of the proposed sequential method in a higher-dimensional input and output setting than the *MOP2* function. Introduced by Deb, Thiele, Laumanns and Zitzler (2005), the version of the *DTLZ2* functions used here have  $d = 4$  inputs and  $m = 4$  outputs. The input space is taken to be  $[0, 1]^4$  and the *DTLZ2* functions are used as the means of the four physical system outputs:

$$\begin{aligned}
 \mu_1^p(\mathbf{x}) &= (1 + g(x_4)) \cos\left(\frac{\pi x_1}{2}\right) \cos\left(\frac{\pi x_1}{2}\right) \cos\left(\frac{\pi x_3}{2}\right) \\
 \mu_2^p(\mathbf{x}) &= (1 + g(x_4)) \cos\left(\frac{\pi x_1}{2}\right) \cos\left(\frac{\pi x_1}{2}\right) \sin\left(\frac{\pi x_3}{2}\right) \\
 \mu_3^p(\mathbf{x}) &= (1 + g(x_4)) \cos\left(\frac{\pi x_1}{2}\right) \sin\left(\frac{\pi x_2}{2}\right) \\
 \mu_4^p(\mathbf{x}) &= (1 + g(x_4)) \sin\left(\frac{\pi x_1}{2}\right), \tag{34}
 \end{aligned}$$

where  $g(x_4) = (x_4 - 0.5)^2$ . The true Pareto Set of  $(\mu_1^p(\mathbf{x}), \mu_2^p(\mathbf{x}), \mu_3^p(\mathbf{x}), \mu_4^p(\mathbf{x}))$  is

$$\mathcal{P}_{\mathcal{X}} = \{\mathbf{x} : x_4 = 0.5\};$$

$\mathcal{P}_{\mathcal{X}}$  is approximated by 1000 points uniformly spread in  $\mathcal{P}_{\mathcal{X}}$ ; the approximation of the true Pareto Front was taken to be the means evaluated at this set of 1000 points.

The physical system observations were generated from  $Y_\ell^p(\mathbf{x}) = \mu_\ell^p(\mathbf{x}) + \epsilon_\ell$ ,  $\ell = 1, 2, 3, 4$ , where  $\epsilon_\ell \sim N(0, .01^2)$ . The simulator runs were taken to have a single calibration input,  $t$ , with domain  $[0, 1]$  and are obtained by adding  $\mu_\ell^p(\mathbf{x}) + (t - 0.5)^2$  to modified versions of  $\mu_\ell^p(\mathbf{x})$  in which  $g(x_4)$  is replaced by  $g^s(x_1, x_4, t) = x_1 + (x_4 - 0.5)^2$ . The true value of  $\phi$  was set equal to 0.5; this quantity minimizes the additive bias, although the simulator is still biased because the multiplicative term  $(1 + g(x_4))$  is perturbed.

While there is difficulty visualizing the augmented PF, the hypervolume indicator plots were simple to visualize. The right panel of Figure 4 plots all HIs based on the reference point  $\mathbf{r} = (1.1, 1.1, 1.1, 1.1)$ . This figure shows that adding the proposed experimental output from the physical system was more effective, in the sense of increasing HI, than adding simulator run data.

## 6 Summary and Discussion

This paper proposes a sequential methodology for determining the Pareto Front and Pareto Set for the mean outputs of a multiple-output physical system measured with error. The proposed method estimates the means of the physical system by a calibrated simulator and adaptively selects new inputs in such a way that the corresponding estimated outputs are spread over the PF. When the sequential procedure takes an additional observation from the physical system, the updated minimax fitness function is used to guide the search for the next control input setting (Sections 4.1–4.3). When the procedure makes an additional run using the computer simulator, the next control input setting is selected as for the physical experiment, while the next calibration parameter setting is chosen to minimize the sum of the posterior mean square prediction errors (Sections 4.2 and 4.4). Section 5 illustrates the efficiency of the method to spread the estimated

outputs on the PF using the 2-dimensional *MOP2* function and the 4-dimensional *DTLZ2* function.

No formal stopping rule has been included in the sequential Pareto minimization methodology in Section 4. Basing such a rule, e.g., on the expected minimax fitness function is possible, although an average of these values over a number of observations is more stable than any single value because this function is not monotone decreasing. The lack of monotonicity is due to the estimated mean response of the physical system being updated after each new observation is taken.

When bias in the simulator is non-additive, we have seen through examples (not shown here) that our proposed approach is more efficient for Pareto minimization than a simulator-only based method. This is perhaps not surprising since the PF of the biased simulator is not necessarily the same as that of the true PF of the means of the corresponding physical system.

For minimizing the integrated mean squared (prediction) error of the physical system output, Ranjan, Lu, Bingham, Reese, Williams, Chou et al. (2011) found that addition of physical observations is more valuable than addition of simulator data. However, as seen in the left panel of Figure 4, this conclusion is not necessarily true for Pareto minimization of the means of the physical system. This leads to the important issue of which experimental platform should be selected at a given stage of the sequential procedure when it is possible to add either an observation from the physical system or a run of the computer simulator.

A possible way to select the experimental platform is first to compute the next control and calibration inputs  $\mathbf{x}^*$  and  $\mathbf{t}^*$ , respectively, as in Section 4.2. Then compare the posterior mean squared prediction errors (MSPE) for outputs from the two experimental platforms given the current data  $\mathcal{Y}^c$  and select the platform producing the *smaller* of these two posterior MSPEs. The posterior MSPEs are calculated as follows. The posterior MSPE of adding the simulator runs  $Y_\ell^s(\mathbf{x}^*, \mathbf{t}^*)$ ,  $\ell = 1, \dots, m$ , is

$$\text{MSPE}_S = \sum_{\ell=1}^m E\{[\hat{U}_\ell^s(\mathbf{x}^*) - U_\ell(\mathbf{x}^*)]^2 \mid \mathcal{Y}^c\}; \quad (35)$$

where  $\hat{U}_\ell^s(\mathbf{x}^*)$  is the posterior mean of  $[U_\ell(\mathbf{x}^*) \mid \mathcal{Y}_\ell, Y_\ell^s(\mathbf{x}^*, \mathbf{t})]$ . The posterior

MSPE of adding one physical observations  $Y_\ell^p(\mathbf{x}^*)$ ,  $\ell = 1, \dots, m$ , is

$$\text{MSPE}_P = \sum_{\ell=1}^m E\{[\hat{U}_\ell^p(\mathbf{x}^*) - U_\ell(\mathbf{x}^*)]^2 \mid \mathcal{Y}^c\}, \quad (36)$$

where  $\hat{U}_\ell^p(\mathbf{x}^*)$  is the posterior mean of  $[U_\ell(\mathbf{x}^*) \mid \mathcal{Y}_\ell, Y_\ell^p(\mathbf{x}^*)]$ . Both expressions depend only the current data  $\mathcal{Y}^c$ . If  $\text{MSPE}_P < \text{MSPE}_S$ , the procedure conducts a physical experiment at  $\mathbf{x}^*$  and adds  $\mathbf{y}^p(\mathbf{x}^*)$ ; if  $\text{MSPE}_S \leq \text{MSPE}_P$  the procedure make a simulator run at  $(\mathbf{x}^*, \mathbf{t}^*)$  and adds  $\mathbf{y}^s(\mathbf{x}^*, \mathbf{t}^*)$ . The performance of this procedure is currently under study.

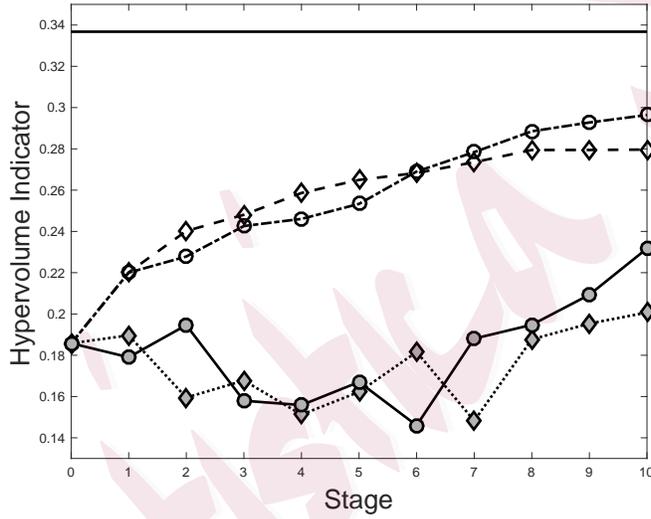


Figure 5: Hypervolume indicators (HIs) of the estimated Pareto Front for the MOP2 example for four procedures: Section 4.3 method adding only simulator runs (hollow diamonds); Section 4.3 method adding only physical system observations (hollow circles); use Mm LHDs of increasing size for extra simulator runs (filled diamonds); use Mm LHDs of increasing size for extra physical system observations (filled circles).

An issue raised by a referee concerns the performance of the single-platform procedures proposed here compared with a mixed-platform procedure in an application where a researcher may wish to implement an alternative *predetermined sequence* of computer and physical system experiments. The need for such an alternative may be due to constraints on the availability of an experimental apparatus or that of a subject-matter specialist to run a (complicated) simulator.

While no comparisons of the two single-platform procedures have been made with such a predetermined mixed-platform procedure, we speculate that, based on the closeness of the HIs shown in Figures 4, the performance of a mixed-platform procedure would be similar to either of the single platform procedures.

A further issue raised by a referee concerns the efficacy of the adaptive procedures compared to that of space-filling, but non-adaptive, designs of the same size. Regarding the HIs of the estimated PFs as a measure of efficiency, Figure 5 plots the HI values for the two adaptive sequential procedures and two sequences of non-adaptive designs for the MOP2 example. All four procedures start with 10 initial inputs for the physical experiment and 15 runs for the computer simulator where the model outputs, measurement errors, and inputs are selected as in Section 5.1. For  $k$  additional observations,  $k = 1, \dots, 10$ , the HIs are denoted by hollow circles for the adaptive procedure that always adds one physical system observation, and by hollow diamonds for the adaptive procedure that always adds one simulator run. Notice that adaptively adding physical experiments seems to have a slight advantage over adaptively adding simulator runs. In contrast, for  $k = 1, \dots, 10$ , each filled circle shows the HI of the estimated PF for a fixed design that uses a  $15 \times 3$  maximin LHD to specify the control and calibration inputs for the simulator runs, and a  $(10 + k) \times 2$  maximin LHD to specify the control inputs for the physical experiment. For  $k = 1, \dots, 10$ , each filled diamond shows the HI for a fixed design that uses a  $10 \times 2$  maximin LHD to specify the control inputs for the physical experiment and a  $(15 + k) \times 3$  maximin LHD to specify the control and calibration inputs for the simulator runs. Figure 5 shows that both adaptive sequential procedures are greatly superior to the fixed size designs.

An issue of current work is the determination of the fraction of observations to be allocated to the initial physical and simulator designs. Intuitively, the size of the initial designs should not be too small, because this can lead to poor estimation. However, if the initial designs use too large a fraction of a budgeted sample size, it may only be possible to identify a few points in the PS, resulting in only a few points around the PF. The optimal initial design sizes may depend on the complexity/smoothness of the output surfaces.

Another issue of current investigation is the extension of our proposed min-

imax fitness function (14) to dependent calibration models for the multiple outputs. When outputs are correlated, a multivariate calibration model could, intuitively, provide more accurate predictions than independent Bayesian calibration models for each output. Currently available multivariate calibration models assume there is no bias between the physical system and corresponding simulator outputs. e.g., Paulo, García-Donato and Palomo (2012) construct a multivariate calibration model using the linear model of coregionalization (Gelfand, Schmidt, Banerjee and Sirmans, 2004). Thus further work is needed.

### Supplementary Material

The supplementary material contains (i) the proof of the bivariate distribution of the calibrated predictor used in Section 4.3, (ii) a proof showing that when the mean function  $\mu(\mathbf{x})$  is known, the minimax fitness function (14) is zero at any  $\mathbf{x}$  that has been previously used in either experimental platform, and (iii) an example showing that the global maximum of the expected improvement surface for the MOP2 function of Section 5.1 is considerably larger than the corresponding values for the 25 points in the initial design.

### Acknowledgements

The authors thank the referees for their helpful suggestions. This work was supported by NSF grants DMS-1310294 and DMS-1564395 (both awarded to The Ohio State University).

### References

- Andrieu, C. and Thoms, J. (2008). A tutorial on adaptive MCMC. *Statistics and Computing* **18**, 343–373.
- Bayarri, M. J., Berger, J. O., Paulo, R., Sacks, J., Cafeo, J. A., Cavendish, J., Lin, C.-H. and Tu, J. (2007). A framework for validation of computer models. *Technometrics* **49**, 138–154.
- Cao, Y. (2009). Pareto front. MATALB File Exchange. <http://ppt.cc/rR3Do>.
- Chen, P.-H. (2016). *Modeling Multivariate Simulator Outputs with Applications to Prediction and Sequential Pareto Minimization*. Ph.D. thesis, Department of Statistics, The Ohio State University, Columbus, Ohio USA.

REFERENCES<sup>24</sup>

- Deb, K., Thiele, L., Laumanns, M. and Zitzler, E. (2005). Scalable Test Problems for Evolutionary Multi-Objective Optimization. A. Abraham, R. Jain and R. Goldberg (Editors), *Evolutionary Multiobjective Optimization: Theoretical Advances and Applications*, chap. 6. Springer, 105–145.
- Finkel, D. E. (2003). DIRECT optimization algorithm user guide. Tech. Rep. crsc-tr03-11, Center for Research in Scientific Computation, North Carolina State University (2003).
- Fonseca, C. M. and Fleming, P. J. (1995). Multiobjective genetic algorithms made easy: Selection sharing and mating restriction. *Genetic Algorithms in Engineering Systems: Innovations and Applications, 1995. GALEZIA. First International Conference on (Conf. Publ. No. 414)*, 45–52.
- Forrester, A., Sobester, A. and Keane, A. (2008). *Engineering design via surrogate modeling: a practical guide*. Wiley, Chichester, UK.
- Gelfand, A., Schmidt, A., Banerjee, S. and Sirmans, C. F. (2004). Nonstationary multivariate process modeling through spatially varying coregionalization. *Sociedad de Estadística e Investigación Operativa Test* **13**, 263–312.
- Han, G., Santner, T. J. and Rawlinson, J. J. (2009). Simultaneous determination of tuning and calibration parameters for computer experiments. *Technometrics* **51**, 464–474.
- Higdon, D., Kennedy, M., Cavendish, J., Cafeo, J. and Ryne, R. (2004). Combining field data and computer simulations for calibration and prediction. *SIAM Journal of Scientific Computing* **26**, 448–466.
- Higdon, D., Gattiker, J., Williams, B. and Rightley, M. (2008). Computer model calibration using high dimensional output. *Journal of the American Statistical Association* **103**, 570–583.
- Jones, D. R., Schonlau, M. and Welch, W. J. (1998). Efficient global optimization of expensive black-box functions. *Journal of Global Optimization* **13**, 455–492.
- Keane, A. J. (2006). Statistical improvement criteria for use in multiobjective design optimization. *AIAA Journal* **44**, 879–891.
- Kennedy, M. C. and O’Hagan, A. (2001). Bayesian calibration of computer models (with discussion). *Journal of the Royal Statistical Society B* **63**, 425–464.
- Leatherman, E. R., Guo, H., Gilbert, S. L., Hutchinson, I. D., Maher, S. A. and Santner, T. J. (2014). Using a statistically calibrated biphasic finite element model of the human knee joint to identify robust designs for a meniscal substitute. *Journal of Biomechanical*

*Engineering* **136**, 071007.

Paulo, R., García-Donato, G. and Palomo, J. (2012). Calibration of computer models with multivariate output. *Computational Statistics & Data Analysis* **56**, 3958–3974.

Picheny, V. (2015). Multiobjective optimization using gaussian process emulators via stepwise uncertainty reduction. *Statistics and Computing* **25**, 1265–1280.

Ranjan, P., Bingham, D. and Michailidis, G. (2008). Sequential experiment design for contour estimation from complex computer codes. *Technometrics* **50**, 527–541.

Ranjan, P., Lu, W., Bingham, D., Reese, S., Williams, B. J., Chou, C.-C., Doss, F., Grosskopf, M. and Holloway, J. P. (2011). Follow-up experimental designs for computer models and physical processes. *Journal of Statistical Theory and Practice* **5**, 119–136.

Sacks, J., Welch, W. J., Mitchell, T. J. and Wynn, H. P. (1989). Design and analysis of computer experiments. *Statistical Science* **4**, 409–423.

Svenson, J. and Santner, T. J. (2016). Multiobjective optimization of expensive-to-evaluate deterministic computer simulator models. *Computational Statistics & Data Analysis* **94**, 250–264.

Tuo, R. and Wu, C. J. (2015). Efficient calibration for imperfect computer models. *Annals of Statistics* To appear.

Williams, B. J., Santner, T. J. and Notz, W. I. (2000). Sequential design of computer experiments to minimize integrated response functions. *Statistica Sinica* **10**, 1133–1152.

Zitzler, E. and Thiele, L. (1999). Multiobjective optimization using evolutionary algorithms - a comparative case study. *Evolutionary Computation, IEEE Transactions on* **3**, 251–271.

Zitzler, E., Knowles, J. and Thiele, L. (2008). Quality Assessment of Pareto Set Approximations. J. Branke, K. Deb, K. Miettinen and R. Slowinski (Editors), *Multiobjective Optimization: Interactive and Evolutionary Approaches*. Springer, 373–404.

Department of Statistics, The Ohio State University, Cockins Hall, 1958 Neil Ave., Columbus, OH 43210, U.S.A.

E-mail: chen.2719@osu.edu

Department of Statistics, The Ohio State University, Cockins Hall, 1958 Neil Ave., Columbus, OH 43210, U.S.A.

E-mail: santner.1@osu.edu

REFERENCES<sup>26</sup>

Department of Statistics, The Ohio State University, Cockins Hall, 1958 Neil Ave., Columbus,  
OH 43210, U.S.A.

E-mail: dean.9@osu.edu

Statistica Sinica